DOCUMENT RESUME

ED 431 849 UD 033 011

AUTHOR Du, Yi; Heistad, Dave
TITLE School Performance Accountability in Minneapolis Public Schools.
PUB DATE 1999-04-00
NOTE 23p.; Paper presented at the Annual Meeting of the American Educational Research Association (Montreal, Quebec, Canada, April 19-23, 1999).
PUB TYPE Reports - Evaluative (142) -- Speeches/Meeting Papers (150)
EDRS PRICE EDRS Price MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Accountability; Analysis of Covariance; Elementary Secondary Education; Public Schools; Reading Achievement; *School Effectiveness; Student Characteristics; *Teacher Effectiveness
IDENTIFIERS Hierarchical Linear Modeling; *Minneapolis Public Schools MN; *Value Added Model

ABSTRACT

This paper describes the accountability system of the Minneapolis Public Schools (Minnesota) and presents the indicators used in the system. It also provides a practical application of two value-added models in the evaluation of effective schools and teachers for this large urban school district. These models are used to rank schools and teachers on their effectiveness at improving student reading achievement and also used to evaluate the factors that contribute to teacher effectiveness. Data are raw normal curve equivalent (NCE) scores from the 1993 and 1994 California Achievement Tests for first and second graders in the Minneapolis schools. The findings from comparing the two models, one an analysis of covariance (ANCOVA) model and the other a hierarchical linear modeling (HLM) approach, show that the results from the two models, such as slope parameters, teachers' value-added values, and schools' value-added values, are nearly identical. Teacher accountability and school accountability can be addressed with the same value-added models. Using the value-added approach and controlling for student characteristics such as poverty, gender, race, special learning needs, and prior achievement, makes it possible to distinguish teacher instructional effects and school effectiveness from externals outside the influence of the teacher and school. (Contains 7 tables, 2 figures, and 12 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Yi Du

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

ED 431 849

School Performance Accountability in Minneapolis Public Schools

Yi Du, Ph.D.
Dave Heistad, Ph.D.

Research, Evaluation and Assessment
Minneapolis Public School

Paper is for the Presentation at the Annual American Educational Research Association Conference in Montreal, Canada, April 1999

BEST COPY AVAILABLE

UD 033011

Increasing accountability for student performance is a main goal for the Minneapolis Public Schools (MPS). This district is building a multifaceted accountability system that encompasses individual, school, and system accountability for improved student learning. At the individual level, student accountability is being forged in the form of curriculum content and performance standards which are aligned with standards-based assessments. The district uses a standardized item bank with multiple test levels for each grade-level to include all students in an accountability system designed to measure growth towards high academic standards as well as current status relative to those standards.

The purpose of this paper is to introduce the Minneapolis Public Schools accountability system. In particular, this paper will show (1) what key indicators are used in the Minneapolis Public Schools accountability system; (2) how to use value-add statistical analyses to calibrate school and teacher performance; (3) how to interpret the empirical evidence and help district and school staff to evaluate school and teacher performance.

The school level accountability system includes six main parts:

- School Performance Continuum
- School Improvement Reports (SIR)
- School Improvement Plans (SIP)
- School Improvement Plan Feedback (SIF)
- Quality Improvement Process (QIP)
- Quality Performance Award (QPA)

These accountability tools are different aspects or facets of a single continuous improvement process at each school site[1]. In order to facilitate this process, the District is developing a common framework and criteria to guide those engaged in feedback to schools. The framework is aligned with the Strategic Direction of the Minneapolis Public Schools, the District Improvement Agenda, curriculum and performance standards for learning, and school quality standards. It will assure accountability consistent with each school's expectations for improvement and with District standards. In addition to use in self-study and planning by schools, it will guide the work of those providing feedback to schools, such as school planning feedback (SIF) , school visits by a quality cadre (QIP), school improvement reports, and monetary performance awards (QPA).

[1] See appendix for a flowchart of the linkages among accountability processes.

Continuous feedback on student performance, attendance, building climate, etc. supports teachers and the site leadership teams to staff openly discuss priorities for allocating resources and adapting instruction to enhance student achievement. The basic data regarding student performance is included in an annual school information report (SIR) provided to principals, teachers, parents and other leadership team members in early August approximately a month prior to the opening of school. These reports include current status in reference to the state high stakes graduation assessments and the Minnesota Basic Standards Tests (MBST). School information reports also provide continuous scale score information on the Northwest Achievement Levels Test which has been equated to the MBST and provides school and grade-level information on the percentage of students "on-course" to pass the MBST in each area of math and reading standards.

The school information report (SIR) also provides math and reading gain score by grade-level referenced to national growth norms for students who are continuously enrolled across two testing periods (e.g., from end of 2nd grade to end of 3rd grade). Recent installation of computer-adaptive versions of the Northwest Achievement Levels Tests allow testing of students new to the district who were not previously included in the accountability for academic progress.

The main component of professional accountability resides in the Professional Development Process (PDP) for teachers, principals, and central office staff. This process includes self-study and coaching relationships in developing and monitoring individual professional development on key goals aligned with the District Improvement Agenda. Recently a study of teacher value-added effects in second grade reading provided feedback on the instructional methods associated with exceptional instruction (Heistad, 1999). Minneapolis has used the value-added analysis to focus attention on teachers "who beat the odds" with an eye towards replicating successful practices rather than using these indicators to identify "low performing" teachers.

A set of statistical models are used to combine "core indicators" that place school, student and teacher performance on a continuum. District Administration use the results of the statistical analysis to inform school feedback ranging from monetary bonuses (QPA) to intensive support and ultimately the possibility of a "fresh start" (i.e. reconstitution) of the school.

Indicators For Identifying Effective Schools and Teachers

The Minneapolis accountability system includes a variety of important educational variables. To measure and monitor school performance, the core combination of indicators includes: (1) Student achievement gain compared to expected gain, (2) Narrowing of gaps in achievement between different groups of students and district and state defined standards, (3) Learning climate, (4) Safety, (5) Student satisfaction and involvement, (6) Family satisfaction and involvement (7) Effective site staff, (8) Staff issues resolved effectively, (9) High community trust. The indicators included in the Minneapolis accountability system are defined in Table 1.

Table 1. School Performance Management Continuum[2]

School Development with Considerable Autonomy	School Needs Some Support	Intensive Assistance Needed
Indicators	Indicators	Indicators
Student achievement gain/growth	Student achievement stagnant	Student achievement in decline
Gap narrowing	Gap steady	Gap widening
High percentage of students engaged (in high standards work)	Medium percentage of students engaged (in high standards work)	Low percentage of students engaged (in high standards work)
Good learning climate: respect responsible actions rigor professionalism	Learning climate signs of trouble: Perceptions of: resistance, blame secrecy, stress, fear	Learning climate very negative (over consecutive weeks) dysfunction denial
school is safe	safety concerns	unsafe
High student attendance and satisfaction	Student absences & suspensions rising	Students transfer/dropout
Families satisfied with school & involved in student success	Pattern of family complaints	Families withdraws or leaves
Effective sit leadership	New site leadership/ developing site leadership	Site leadership incompetence/isolation
Effective site staff	Ineffective or new site staff	Incompetent site staff
Staff issues resolved effectively	pattern of staff complaints	Staff transfer/leave Recurring staff complaints
Community trust high & improving	Community trust unsure	No community trust

[2] This school performance management continuum table is included in the contract with MPS principals and teachers bargaining units along with provision for rewards and intensive assistance.

BEST COPY AVAILABLE

The school performance data on this School Performance Management Continuum are collected from district-wide student achievement assessments; teacher, student and parent surveys; as well as school reports on student suspensions and student attendance rate. The first order indicators come from test scores. A school with improving academic achievement results is viewed as "on course" and free to continue school development with little administrative constraints. A school with decreasing academic performance in one or more areas will be given support and intensive assistance as needed. For example, a school with low and declining achievement in reading, but not math will be given specific assistance in areas of literacy. A school with low and declining achievement in reading and math along with declining indices of learning climate, leadership, and staff and community satisfaction may receive immediate intensive assistance which may take the form of a "fresh start" in the extreme case.

At the district level, the system sets and measures progress on continuous achievement goals. Both the absolute level of performance relative to district and state standards and the value-added contribution of schools to student achievement are reported. Table 2 shows these measurable indicators for the MPS accountability system.

Table 2 Measurable Indicators Used for the Accountability System

Category	Variables	Grades Used
Achievement	Oral Reading	Grades 1, 2
	MCA, reading, math, writing	Grades 3 and 5
	NALT, reading and math	Grades 2 through 7
	MBST, reading and math	Grades 8, 9, 10,& 11
	MBST, writing	Grade 10
	PLAN, reading, English, math, science	Grade 10
	ACT and/or SAT	Grades 11 and 12
Survey	Learning climate	Grades 3, 6, 10
	Safety	
	Student satisfaction and involvement	
	Family Satisfaction and Involvement	
	Effective site staff	
	Staff issues resolved effectively	
School Outcomes	Attendance	All grades combined
	Promotion Rate	
	Dropout rate	

Notes:
NALT = Northwest Achievement Level Tests
MCA = Minnesota Comprehensive Assessment
MBST = Minnesota Basic Standards Tests (required for graduation)

Performance Awards and Sanctions

All results and findings from value-added analysis and other statistical analyses are described in MPS accountability system reports. Findings of these reports are provided bi-annually for all schools in Minneapolis. The reports include information on student achievement for each subject and grade for current year, such as percentage of students who are on-course to pass the MBST tests, percentage of students who met the MPS district curriculum standards, Percentage of students who had one-year or above gains, comparison of racial/ethnic group gains; information on value-added analysis; as well as information from survey for teachers, students and parents.

The information supplied in the reports has been valued by school and district administrators involved in the School Improvement Plans, the school improvement feedback, curricular planning, program evaluation, and teacher developing strategies to meet the needs of students with differing academic abilities. These reports allow the school and district to pinpoint grades or subject area strengths and weaknesses as well as to direct efforts and resources accordingly.

Cash awards, know as qualify performance awards (QPA), were provided to Minneapolis Schools for the first time in 1997. Schools applied for the awards based on achievement growth data, attendance, advanced level course participation, and engagement in individualized staff development planning. At the end of the 1998-99 school year monetary awards will again be available for schools which demonstrate exceptional performance. Awards will be funded from a $400,000 grant and should enable school who qualify to receive between $10,000 and $20,000 per school. Awards will be based in part on the results of value-added analysis from one of the following models.

Value-Added Models and School Accountability

In order to make schools and teachers accountable, appropriate statistical models need to be used. In Minneapolis Public Schools, "value-added" statistical analyses are applied to determine school and teacher effectiveness. The general concept of the value-added analysis is to compare a student's observed test score to the prediction of how well the student can be expected to do, based on his or her performance during the previous year. An estimate of the deviation from expectation , the student-level residual, is considered to be the "value-added" by the teacher or, at the more aggregate level, by the school.

The advantage of a value-added approach to the measurement of growth/gain have been demonstrated in recent years (Bryk, Deabster, Easton, Luppescu, & Thum, 1995; Meyer, 1993 & 1996). As Meyer (1996) has demonstrated, significant school improvements can be occurring but are masked under status-based accountability systems, especially under conditions of high student mobility and other factors that depress average scores.

Raudenbush and Willms (1995) and Meyer (1996) distinguished between *Type A* and *Type B* school effects. Type A effects, which Meyer calls "total performance indicators", are often the interest of parents and real estate agents, while Type B effects, which Meyer calls "intrinsic indicators", are of more interest to education policy makers and evaluators. Type A effect is the total performance which includes school's effectiveness derived from the practice of its staff and favorable student composition, and from the beneficial influence of the social and economic context of the community in which the school is located. While type B effect is the intrinsic value-added effect of school practice on student learning unconfounded by school context variables. The value-added models are ideally suited to estimate Type B effects because they provide an index of school practice variables (curriculum content, instructional practice, and school resources) after factoring out the influence of school context variables (student demographics, community characteristics). "The Type B effect is the effect school officials consider when evaluating the performance of those who work in the schools. A school with an unfavorable context could produce a large Type B effect through the effort and talent of its staff. The school would rightly earn the respect of school evaluators even through parents shopping for a large Type A effect might not want to choose that school" (p. 310).

Using the value-added approaches, two statistical models, a single-stage ANCOVA model where student characteristics and initial achievement level are controlled and a two-stage hierarchical linear model where individual and school characteristics are controlled statistically, are employed to estimate value-added effects for schools and teachers. These analyses were designed to isolate teacher effects and school performance from other sources of achievement variance so that instructional variables and school performance variables associated with student achievement and gain would be identified while controlling statistically for student characteristic difference.

This study will apply the value-added models in examining Type B school and teacher effects. In particular, a value-added school effective index for this district will be obtained controlling for student family socio-economic status, school population's percent of student poverty, and so on. The analyses progress from the simple to more complex in order to determine the model best suited to use in the Minneapolis School Performance Continuum and Quality Performance Awards (QPA).

The single-stage value-added regression models are defined as the follows:

$$\mathbf{Y_{is}} = \gamma + \theta\, \mathbf{X_{is}} + \alpha_{\mathbf{is}} + \eta_{\mathbf{s}} + \varepsilon_{\mathbf{is}} \qquad \textbf{(1)}$$

where
Y_{is} represents student post test scores;
i indexes individual students;
s indexes teachers or schools;
γ is a constant;
X_{is} represents student pretest scores;
α_{is} represents student characteristics including free or reduced price lunch, resides with two or single parent(s), limit English Proficient, Special education, race and so on;
θ represents model parameter for pretest score;
α represents model parameter for student characteristics;
η represents teacher effect or school effect after controlling for all student-level factors;
$\varepsilon_{is} \sim N(0, \sigma^2)$, is the error term.

The hierarchical linear models have two levels as:

Level 1: $$Y_{ij} = \beta_{0j} + \beta_{1j} X_{1j} + \beta_{2j} X_{2j} + \ldots + \beta_{kj} X_{kj} + r_{ij} \qquad (2)$$

Level 2:
$$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11} W_j + u_{1j}$$
$$\beta_{kj} = \gamma_{k0} + \gamma_{k1} W_j + u_{kj}$$

At Level 1, student level,
Y_{ij} is dependent variables and represent student achievement;
β_{01} is intercept, and defined as the expected student achievement;
β_{k0} is the average achievement slope associated with student characteristics;
r_{ij} i~ $N(0, \sigma^2)$, is the error term.

At Level 2, school level
β_{kj} becomes independent variables;
γ_{00} is the average school mean for the population of schools;
γ_{01} is the mean achievement difference between different schools;

γ_{10} is the average achievement slope associated with school characteristics;
γ_{11} is the mean difference in slopes between different schools;
u_{1j} is the unique effect of school j in the slope holding W_j constant;
u_{0j} is the unique effect of school j on mean achievement holding W_j constant.

Data and Analysis

The raw data used in this study are NCE scores from the 1993 and 1994 California Achievement Tests (CAT). All MPS students in Grades 2 through 8 took CAT in reading and mathematics before 1996. This data set includes all first grade and second grade students' CAT reading scores during 1993 and 1994. All first grade students received the Level 11 test and all second grade students received the Level 2 test. About 34 reading comprehension items for Level 11 and 38 reading comprehension items for Level 12. The analysis includes students who stayed in the same school for two years for the school to impact their education as well as for all students eligible for the testing programs. For the study, student NCE scores (norm curve equivalence) were used and also linked with information on student's demographic data and school characteristics data. Using these data, an example of applying these models to determine school and teacher effectiveness is provided in this paper.

One-Stage ANCOVA Model

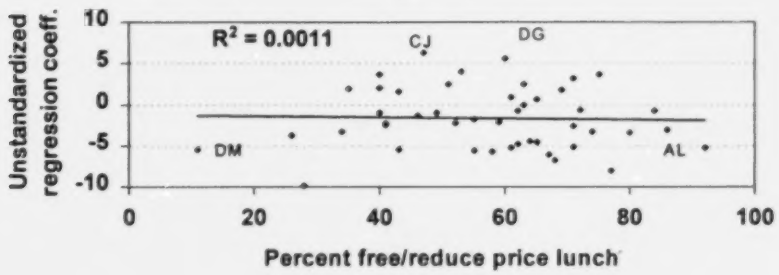
Using the 1993-94 data and SPSS® (version 8.0 for windows), the one-stage regression model was run with a constant, and Total Reading NCE scores for the pretest and posttest. All demographic factors and school effects were coded dichotomously (0,1) and entered simultaneously with the 1993 Total Reading NCEs as predictor's in the model. The 1994 Total Reading score was the dependent variable in this analysis. Unstandardized regression coefficients and standard errors are presented in Table 3. Using this approach all teacher effects appear in the SPSS output along with their respective standard errors.

Table 3. Unstandardardized Regression Coefficients for One-Stage Model

	Coefficient	Std. Error	t ratio	p value
(Constant)	20.85	1.72	12.13	0.00
READ93	0.73	0.01	55.98	0.00
HIPOVRTY	-0.37	0.61	-0.61	0.54
BOTHPAR	0.37	0.60	0.62	0.54
FRLUNCH	-4.58	0.67	-6.88	0.00
SPED	-4.64	0.89	-5.23	0.00
AMERIND	-3.66	1.10	-3.32	0.00
AFRAMER	-4.50	0.65	-6.90	0.00
SEX	-1.90	0.50	-3.79	0.00
LEP	-6.41	1.18	-5.45	0.00
school DV	-9.87	3.86	-2.56	0.01
school AK	-3.77	2.33	-1.62	0.11
school BU	-5.23	2.21	-2.37	0.02
school DM	-5.46	3.85	-1.42	0.16
school AL	-5.29	2.11	-2.51	0.01
school BJ	-8.07	3.07	-2.63	0.01
school CO	-5.56	2.00	-2.77	0.01
school CE	-5.13	2.30	-2.23	0.03
school CL	-6.74	2.11	-3.19	0.00
school CB	-2.20	2.34	-0.94	0.35
school AF	-6.11	2.28	-2.69	0.01
school BM	-5.69	2.10	-2.71	0.01
school DQ	-5.51	2.52	-2.19	0.03
school CV	-4.45	2.13	-2.09	0.04
school DR	-3.40	2.48	-1.37	0.17
school DJ	-3.27	2.10	-1.56	0.12
school BO	-4.85	2.22	-2.18	0.03
school BS	-2.58	2.15	-1.20	0.23
school BK	-4.58	2.32	-1.97	0.05
school BE	-3.35	2.27	-1.48	0.14
school BD	-2.53	1.77	-1.43	0.15
school DX	-2.15	4.41	-0.49	0.63
school CH	-3.08	1.74	-1.77	0.08
school AI	-2.25	2.61	-0.86	0.39
school BB	-1.37	2.16	-0.63	0.53
school AY	-1.79	1.91	-0.94	0.35
school CZ	-0.74	2.52	-0.29	0.77
school DC	-0.01	2.36	0.00	1.00
school CT	-0.98	2.01	-0.49	0.63
school AT	-0.80	2.26	-0.36	0.72
school BI	1.80	2.13	0.84	0.40
school DI	1.54	2.61	0.59	0.56
school DB	0.60	2.28	0.26	0.79
school AG	-1.00	1.98	-0.51	0.61
school DZ	0.86	2.23	0.38	0.70
school AW	1.88	2.42	0.78	0.44
school DY	2.06	2.24	0.92	0.36
school CR	3.59	3.02	1.19	0.23
school BF	-0.58	2.31	-0.25	0.80
school BV	2.45	2.02	1.21	0.23
school AP	3.67	2.18	1.68	0.09
school CW	3.19	2.20	1.45	0.15
school CJ	6.28	2.27	2.77	0.01
school DH	2.45	2.18	1.12	0.26
school BQ	3.95	2.21	1.79	0.07
school DG	5.57	2.42	2.30	0.02

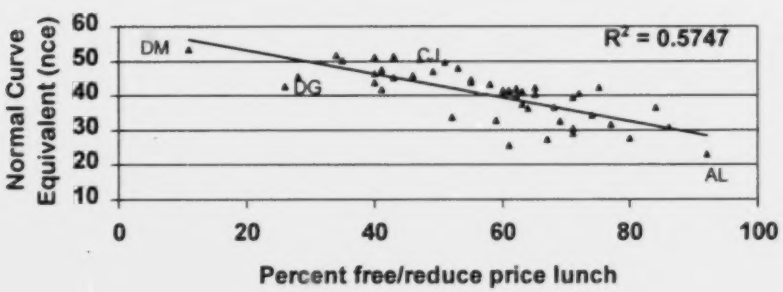
The value-added coefficients from Table 3 are graphed along with corresponding school level information on the percentage of students receiving free or reduced price lunch. The R^2 (.001) indicates no linear relationship between poverty and value-added for this analysis.

Figure 1. Grade 2 Reading Value-added by free or reduced price lunch percentage



This graph may be compared to the scatterplot indicating the correlation between poverty and average test scores (see Figure 2) using the post-test alone as an indicator of school success. School DM had the highest overall test score at 53.4 nces (i.e. about the 56th percentile) and school AL had the lowest average test score at 23 nces (i.e. about the 10th percentile).[1] This does not mean, however, that school AL produced the greatest reading value-added. In fact, you can see from Figure 1, that school DM and school AL had equivalent value-added estimates at about -5 nces. School CJ and DG actually produced the highest value-added even though their average test scores were in the middle range.

Figure 2. Grade 2 Reading Average Test Score by free or reduced price lunch %



[1] DM had the lowest average poverty at 11% and AL had the highest level poverty at 92%. Poverty predicted 57% of the variance in average test scores.

HLM Procedure

Using the data and HLM2L programs (Bryk, Randenbush, and Congdon, 1996), the different models, one-way ANOVA, random-coefficient model, and intercept- and slope-as-outcome model were tried as follows:

The One-Way ANOVA model

Level 1: $Y_{ij} = \beta_{oj} + \gamma_{ij}$

Level 2: $\beta_{oj} = \gamma_{oo} + u_{oj}$

The Random-Coefficient model:

Level 1: $Y_{ij} = \beta_{oj} + \beta_{oj} (X_{ij} - \bar{X}_{.j}) + \gamma_{ij}$

Level 2: $\beta_{oj} = \gamma_{oo} + u_{oj}$

$\beta_{kj} = \beta_{kj} + u_{oj}$

The Intercept- and Slope-as-Outcomes model:

Level 1: $Y_{ij} = \beta_{oj} + \beta_{oj} (X_{ij} - \bar{X}_{.j}) + \gamma_{ij}$

Level 2: $\beta_{oj} = \gamma_{oo} + \gamma_{o1} w_j + u_{oj}$

$\beta_{kj} = \gamma_{ko} + \gamma_{k1} w_j + u_{kj}$

By trying different models above, a random-intercept model was selected as the best model shown in equation (3) for the study. The best model will provide effect analysis to show which factors affect school performance and a value-added school effects to determine which schools "beat the odds." Combining different models together, the variance components analysis can describe how much variance is explained at both school and student levels.

Level 1 or student level

$$Y_{ij} = \beta_{oj} + \beta_{1j} (X_{1ij} - X_{1..}) + \beta_{2j} (X_{2ij} - X_{2..}) + \beta_{3j} (X_{3ij} - X_{3..}) + \beta_{4j} (X_{4ij} - X_{4..}) + \beta_{5j} (X_{5ij} - X_{5..}) + \beta_{6j} (X_{6ij} - X_{6..}) + \beta_{7j} (X_{7ij} - X_{7..}) + r_{ij} \quad (3)$$

Where Y_{ij} = CAT 1994 NCE score for student I for school j,

β_{oj} = expected CAT NCE score for a student whose value on X_{1ij} is equal to the grand mean,

β_{1j} = expected change in CAT score for a unit change in student prior score,

β_{2j} = expected in CAT score for a unit change in student gender (male =1 and female=0),

β_{3j} = expected in CAT score for a unit change in student in LEP program (LEP =1 and non LEP=0),

β_{4j} = expected in CAT score for a unit change in student in Special Educational program (Spe. Ed. =1 and non Spe. Ed.=0),

β_{5j} = expected in CAT score for a unit change in student in SES (student with free or reduced lunch =1 and non-free and reduced lunch=0),

β_{6j} = expected in CAT score for a unit change in American Indian student (American Indian =1 and non American Indian=0), and

β_{7j}= expected in CAT score for a unit change in African American student (African American = 1 and non-African American=0).

Level 2 or school level

$\beta_{0j} = \gamma_{00} + \gamma_{01} W_{1j} + \gamma_{02} W_{2j} + u_{oj}$

$\beta_{1j} = \gamma_{10}$

$\beta_{2j} = \gamma_{20}$

.....

$\beta_{7j} = \gamma_{70}$

Where γ_{00} = expected CAT mean for a school,

γ_{01} = the relationship between the expected school mean achievement (β_{0j}) and percent of poverty of the school (W_{1j}) after controlling for other school level variables,

γ_{02} = the relationship between the expected school mean achievement (β_{0j}) and the school prior score mean (W_{2j}) after controlling for other school level variables,

γ_{10} = the fixed value of the slope (β_{1j}) across all schools (pooled within-school regression coefficient),

u_{oj} = unique effect of school j on the average achievement after controlling for W_{1j} and W_{2j}.

In this model, we constrain the Level-2 coefficients γ_{11} ... γ_{j1} and the random effects equal to 0. The resulting model would be an ANCOVA with random effect in intercept and level 1 predictors as covariates.

Table 4 displays the results. At school level, school MEAN SES (percentage of students with free or reduced price lunch) is significantly related to school mean reading achievement, γ_{01} =-0.12, t=-2.81. Also, school PRIOR MEAN SCORE is significantly related to school mean reading achievement, controlling for the effect of MEAN SES, γ_{02} =-0.24, t =-2.493.

Comparing the school mean reading achievement estimate, -0.12, with the student SES slope predictor -4.70, the result indicates that individual students' poverty family backgrounds have more significant impact on student achievement than school composition (Mean SES) does. The prior mean score, -0.24, is beyond our hypothesis that schools with high prior scores have high achievement and verse visa. It may imply that schools have different growth rates and that schools with higher prior scores have less room to make progress than schools with low prior scores. The negative coefficient on the prior score parameter needs further investigation and study.

Combining the results with the one-way ANOVA model, we are able to get variance components analysis at both school level and student level. About 85% of the parameter variation in mean achievement has been explained by MEAN SES and PRIOR MEAN SCORES. A test of null hypothesis that no residual variance remains to be explained, is rejected. The $\chi 2$ statistic is 153.89 with df = j-1 = 44, $p < .001$. Hence, the results encourage a search for further school-level variables that might help account for the remaining variation in the intercepts.

At the student level, these factors, student prior reading score, student SES (as measured by free or reduced price lunch eligibility), gender, LEP or Special educational background, American Indian or African American background significantly affect achievement. Variance component analysis shows that 66% of variance of level 1 parameter variation in mean achievement has been explained by these seven factors. In addition, these analysis shows that the reliability index for the sample mean is .70.

Table 4 Primary Finding of School Effects for HLM Model

Fixed effect	Coefficient	se	t-ratio	P-value
Model for school means				
Intercept, γ_{00}	57.85	6.036	9.582	0.000
Mean SES, γ_{01}	-0.12	0.042	-2.801	0.008
Prior Mean Score, γ_{02}	-0.24	0.10	-2.493	0.017
Model for Prior Score				
slope intercept, γ_{10}	0.73	0.01	56.251	0.000
Model for Male student				
slope intercept, γ_{20}	-1.93	0.50	-3.858	0.001
Model for LEP student				
slope intercept, γ_{30}	-6.37	1.15	-5.531	0.000
Model for Special Ed. student				
slope intercept, γ_{40}	-4.67	0.88	-5.29	0.000
Model for student SES slope				
intercept, γ_{50}	-4.70	0.64	-7.34	0.000
Model for Amer. Indian student				
slope intercept, γ_{60}	-3.99	1.08	-3.71	0.001
Model for Afr. American student				
slope intercept, γ_{70}	-4.53	0.63	-7.22	0.000

Final estimate of variance components:

Random Effect	Variance	df	Chi-square	P-value
Intercept, u_{7j}	7.67	44	153.89	0.000
Level-1, R_{70}	191.12			

An important by-product of the HLM analysis is the residual file that can be used to derive in index of the Type B effectiveness of schools at raising academic achievement after controlling for relevant student and school level variables. Once the index of effectiveness is obtained then schools can be ranked according to this index. The current index controls for the seven factors at student level and two factors at school level. The factors to be controlled can be redefined based on district policy makers' or evaluators' decision.

The value-added effectiveness index is a measure of the school's level of academic achievement after controlling for student and school context variables (in this case seven variables at student level and two variables at school level) as equation (4) shows.

$$\mu_{0j} = Y_{.j} - [\gamma_{00} + \gamma_{01}\,(\text{MEAN SES}) + \gamma_{02}\,(\text{MEAN PRIOR SCORE}) + \gamma_{10}\,(\text{PRIOR SCORE}) + \gamma_{20}\,(\text{GENDER}) + \gamma_{30}\,(\text{LEP}) + \gamma_{40}\,(\text{SPECIAL ED.}) + \gamma_{50}\,(\text{SES}) + \gamma_{60}\,(\text{AMERICAN INDIAN}) + \gamma_{70}\,(\text{AFRICAN AMERICAN})] \quad (4)$$

It can be interpreted as the difference between the school's actual mean performance and the schools expected mean performance after controlling these factors included in equation (4). Table 5 lists the value-added effect index that was calculated for all 45 schools. Criteria can be set up to divide schools or to determine if schools are effective or not. For example, Prince George's County Public Schools (Phillips, G.W. and Adcock, E.P. 1996) used one standard deviation. Schools that are more than one standard deviation above what is expected (based on student SES, prior scores, and so on.) are considered effective. Schools that are within one standard deviation are considered doing about as well as can be expected (no value-added). Schools that are one standard deviation below are considered ineffective and are not performing up to the levels of other schools with similar levels of poverty. In Minneapolis Public Schools, outlier analysis was used. Schools that can be identified as positive outliers are considered "beating the odds," e.g., Schools DG and Schools BQ.

In addition, comparison between value-added effect and school achievement growth, the difference between pre and post achievement scores, is provided in Table 5. This table shows that schools with pure positive growth index do not necessarily mean effective. For example, School BI with 9.32 nce growth from 1993 to 1994. After controlling student characteristics and school contexts, this school's value-added effect is only 1.2, which means this school has some advantages in school contexts or student characteristics. It would be unfair to reward school staff purely on the basis of their achievement without considering school context and student composition and characteristics.

Table 5 Value-Added Index for School Performance

SCHOOL	Value-Add Index	Growth*	Gender	LEP	Spe. Ed.	Indian American	African American	SES	Prior Score	Post Score
DV	-7.7	-9.94	53.3	.0	26.7	6.7	46.7	33.3	55.41	45.48
AK	-6.6	2.80	47.4	.0	17.5	1.8	21.1	24.6	39.24	42.04
BU	-5.1	.83	60.3	.0	17.6	14.7	41.2	75.0	26.18	27.01
DM	-4.5	-3.05	33.3	.0	6.7	.0	26.7	20.0	57.31	54.27
AL	-4.1	-.21	55.6	3.7	13.6	2.5	74.1	91.4	23.11	22.90
BJ	-4.0	-3.81	65.4	.0	3.8	11.5	34.6	96.2	32.46	28.64
CO	-4.0	-1.59	57.3	6.3	3.1	5.2	12.5	44.8	46.91	45.32
CE	-3.4	-.41	45.8	.0	6.8	3.4	57.6	78.0	32.02	31.60
CL	-3.1	-4.45	50.0	24.4	7.7	3.8	24.4	73.1	41.65	37.20
CB	-3.0	3.74	50.0	.0	7.1	7.1	16.1	39.3	40.53	44.27
AF	-2.9	-2.54	54.8	.0	17.7	24.2	43.5	88.7	31.86	29.32
BM	-2.8	-2.77	50.6	5.1	5.1	.0	32.9	58.2	45.91	43.14
DQ	-2.6	-3.88	73.3	.0	15.6	22.2	22.2	53.3	48.11	44.23
CV	-2.1	.16	40.8	9.2	6.6	7.9	26.3	72.4	36.71	36.87
DR	-2.0	2.73	42.3	.0	7.7	65.4	5.8	84.6	26.15	28.88
DJ	-1.8	-.85	54.3	.0	3.7	9.9	28.4	32.1	52.91	52.06
BO	-1.5	-3.31	59.1	19.7	9.1	10.6	25.8	63.6	45.02	41.71
BS	-1.3	2.63	51.4	10.8	2.7	1.4	60.8	82.4	27.38	30.01
BK	-0.9	-2.44	59.6	.0	12.3	8.8	36.8	70.2	43.64	41.21
BE	-0.8	.19	56.5	.0	4.8	4.8	62.9	79.0	34.71	34.89
BD	-0.8	.22	49.4	15.9	8.8	.6	21.8	45.9	47.45	47.67
DX	-0.6	1.34	36.4	.0	18.2	9.1	72.7	72.7	33.25	34.59
CH	-0.6	1.12	44.8	7.3	7.3	6.3	59.4	87.5	30.08	31.20
AI	-0.3	3.06	57.5	.0	7.5	12.5	45.0	85.0	28.85	31.92
BB	-0.2	2.20	52.1	24.7	11.0	1.4	8.2	47.9	43.77	45.97
AY	0.3	1.99	49.6	11.3	5.2	5.2	24.3	59.1	41.98	43.97
CZ	0.3	5.19	43.2	6.8	11.4	.0	15.9	63.6	35.78	40.97
DC	0.4	5.16	44.4	.0	14.8	13.0	38.9	61.1	34.17	39.32
CT	0.8	2.42	40.0	6.3	8.4	3.2	18.9	44.2	48.41	50.83
AT	0.9	4.22	39.7	.0	6.3	20.6	44.4	77.8	31.46	35.68
BI	1.2	9.32	49.4	.0	10.4	15.6	39.0	75.3	23.08	32.40
DI	1.2	5.25	40.0	20.0	10.0	.0	10.0	35.0	47.40	52.65
DB	1.5	6.30	43.3	6.7	5.0	1.7	21.7	60.0	37.03	43.33
AG	1.8	.85	59.8	25.5	9.8	1.0	20.6	58.8	45.32	46.17
DZ	1.9	5.40	56.9	.0	4.6	7.7	49.2	63.1	36.01	41.41
BP	2.0	3.58	58.3	11.5	6.3	1.0	34.4	63.5	39.84	43.42
AW	2.1	6.45	52.0	.0	8.0	14.0	20.0	40.0	43.81	50.26
DY	3.0	7.24	40.6	.0	9.4	3.1	34.4	57.8	38.17	45.41
CR	3.1	10.15	48.1	.0	14.8	.0	33.3	59.3	31.78	41.93
BF	3.4	1.74	41.4	15.5	20.7	6.9	29.3	81.0	39.29	41.02
BV	3.8	6.54	52.7	.0	18.3	8.6	44.1	65.6	35.90	42.44
AP	4.0	8.26	37.1	.0	10.0	2.9	40.0	45.7	41.57	49.83
CW	4.0	8.27	47.8	.0	26.1	4.3	37.7	68.1	32.57	40.84
CJ	4.7	12.82	50.0	.0	1.6	3.2	22.6	35.5	39.25	52.07
DH	5.9	4.13	55.7	.0	8.6	1.4	52.9	61.4	46.65	50.77
BQ	6.2	6.74	46.3	.0	11.9	1.5	53.7	61.2	42.24	48.99
DG	7.5	9.24	64.0	.0	6.0	6.0	66.0	76.0	33.49	42.73

* Growth represents the difference between before and after controlling the variables at student and school levels.

BEST COPY AVAILABLE

Comparison Between the One-Stage ANCOVA and HLM Models

An additional study was conducted to compare the effectiveness and appropriateness of the two statistic models. The same data set and variables were used to analyze both teachers' effects and school effects. As a result, both teacher value-added effect index and school value-added effect index were generated. First, the one - stage ANCOVA model and the fixed-coefficient HLM model were applied for the teacher effect analysis. The comparison between the two model shows that the slope parameters are nearly identical as Table 6 shows.

Table 6 One Stage ANCOVA Model and HLM Comparison for Teacher Effect

Variable	One-Step Model		HLM Model	
	Coefficient	SE	Coefficient	SE
Constant	16.14	1.15	41.4	0.39
Pre-test score	0.73	0.01	0.73	0.01
African American	-4.37	0.66	-4.33	0.64
American Indian	-3.73	1.10	-4.01	1.06
Male	-1.92	0.50	-2.02	0.49
Live with 2 Parents	0.80	0.61	0.69	0.59
Free or reduced price lunch	-4.07	0.67	-4.46	0.65
Resides in high poverty zip code	-0.67	0.62	-0.36	0.60
Limited English Proficiency	-6.59	1.17	-6.91	1.14
Special Education	-3.69	0.89	-4.07	0.88

In addition, the study compared results for school effects from two models. The one-stage ANCOVA model and the intercept-random HLM model were applied into school effect analysis. The two models used almost same predictors except that the HLM model excluded two factors: parents and high poverty areas by zip code because of quite weak effects. Table 7 shows that the common slope parameter estimations are also nearly identical.

BEST COPY AVAILABLE

Table 7 One Stage ANCOVA Model and HLM Comparison for School Effect

Variable	One-Stage Model		HLM Model	
	Coefficient	SE	Coefficient	SE
Prior Scores	0.73	0.01	0.73	0.01
African American	-4.50	0.65	-4.53	0.63
American Indian	-3.66	1.10	-3.99	1.08
Male	-1.90	0.50	-1.93	0.50
Free or reduced price lunch	-4.58	0.67	-4.70	0.64
Limited English Proficiency	-6.41	1.18	-6.41	1.15
Special Education	-4.64	0.89	-4.67	0.88

A correlation analysis was conducted for both teacher value-added effects and school value-added effects generated from both models. The correlation coefficients generated for the teacher value-added indexes was .96 and for school value-added indexes was .92, indicating that there is little difference identifying value-added teachers and schools using the two models.

The identical slope parameters and value-added indexes between the two models are from the similar model definition. In teacher effect analysis, the two models are quite identical when we use the Random-Coefficient HLM without defining any predicators at level-2. In school effect analysis, a random-intercept HLM model was selected. Combining the two-level random-intercept HLM model as equation (5), the only difference between HLM model and one-stage ANCOVA model is that the group effect here, U_{0j}, is conceived as random rather than fixed. Therefore, it would not be surprising that the results from two models are so identical.

$$Y_{ij} = \gamma_{00} + \gamma_{01} W_j + \gamma_{02} W_{2j} + \gamma_{10} (X_{1j} - X_{1.}) + \ldots + \gamma_{7j} (X_{7j} - X_{7.}) + U_{0j} + r_{ij} \qquad (5)$$

Findings and Further Discussion

This paper describes Minneapolis Public Schools accountability system and presents the indicators used in the system. Also, this paper provides a practical application of two value-added models in evaluation of effective schools and teachers for a large urban school district (Minneapolis Public Schools in Minnesota). These models are used to rank schools and teachers on their effectiveness at improving

student reading achievement and also used to evaluate which factors contribute to school or teachers effectiveness.

At the school level, school poverty and school average prior achievement are impacts on school mean achievement. At the student level, student prior test score, student racial/ethnic background, student's special educational needs (i.e., LEP and Special Ed.) and student gender have large impacts on student learning. The parameter estimations show that student level predictors have bigger impacts on student achievement than school level predictors.

The findings from comparing the one-stage ANCOVA model to the HLM model show that the results from the two models such as slope parameters, teachers' and schools' value-added values are nearly identical. This finding is consistent with previous study (Mundlak, 1978). In this study, the HLM model used for teacher effect analysis did not specify any level-2 predicators. As a result, this model is quite identical to the one-stage ANCOVA model. In school effect analysis, the only difference between the two models is that HLM model has the random unique group effect at level-2. The one-stage ANCOVA model produces the total performance indicator, type A effects, and the HLM model with school-level regressors in the level-2 produces the intrinsic performance indictor, type B effects. Both models conduct value-added analysis well in selecting submodels, defining parameters, and generating value-added effects for groups.

An important task in the development of the accountability system is the investigation of the correlates to school and teacher effectiveness and identify other factors that impact teacher and school effectiveness. The findings from initial correlational studies of school effects in Minneapolis found that three factors, attendance, earlier intervention and mobility, play the most important roles student learning on the Minnesota Basic Standards Tests (Myers, 1997). Future plans for value-added study will include examining the impacts of attendance, mobility and parent education on rates of student learning.

Teacher accountability and school accountability can be addressed with the same value-added models. When trying different variables for teachers and schools effect in this study, it is found that not all variables that affect student learning at school level also have large impacts at the teacher level. Further research may examine different variables for teachers and schools that impact student learning. Since students are

nested with teachers (classrooms), teachers are nested with schools, a three level of HLM analysis may be considered as a future area of study.

Using the value-added approach and controlling for student characteristics such as poverty, gender, race, special learning needs and prior achievement, we can distinguish teacher's instructional effects and school effectiveness from external factors outside the influence of the teacher and the school. Hopefully, this paper will be useful to practitioners or researchers interested in methodologies for identifying and combining school performance indicators to make high stakes accountability decisions.

What we have learn from this study will inform decisions regarding the Quality Performance Awards at the end of this school year and enable the Minneapolis Public School District to recognize teachers & schools who truly "beat the odds".

References

Bryk, A. and Raudenbush, S.W. (1992). *Hierarchical Linear Models, Applications and Data Analysis Methods*, Newbury Park: Sage Publications.

Bryk, A. S., Deabster, P., Easton, J.Q., Luppescu, S., & Thum, Y.M. (1995). Measuring achievement gains in the Chicago Public Schools. *Education and Urban Society*, 26, 306-319.

Bryk, A. S., Raudenbush, S., & Congdon, R. (1996). *Hierarchical Linear and Nonlinear Modeling with the HLM/2L and HLM/3L Programs*. SSI Scientific Software International : Chicago.

Heistad, D.J. (1999). *Stability and correlates of teacher effects in grade two reading achievement*. Unpublished Doctorate Dissertation, University of Minnesota.

Meyer, R. H. (1993). *Can schools be held accountable for good performance? A critique of common educational performance indicators*. Working paper, Harris Graduate School of Public Policy Studies, University of Chicago.

Meyer, R. H. (1996). "Value-Added Indicators of School Performance," in Hanushek, Erica and Jorgensen, Dale W. (Eds.), *Improving the Performance of America's Schools: The Role of Incentives*, Washington, D.C.: National Academy Press, pp. 171-196.

Myers, S.L. (1997). *Analysis of the 1996 Minnesota Basic Standards Test Data*. Unpublished Manuscript. Roy Wilkins Center in the Hubert H. Humphrey Institute of Public Affairs. University of Minnesota.

Raudenbush, S.W., & Willms, J.D. (1995). The estimation of School Effects, *Journal of Educational and Behavioral Statistics*, Winter, 20, 4, pp. 307-335.

Millman, J. (1997). *Grading teachers, grading schools: Is student achievement a valid evaluation measure*. Thousand Oaks, CA: Corwin.

Minneapolis Public Schools, (1997 - 1999). *Teacher Contract, Agreements and Policies*, Minneapolis Federation of Teachers, #59.

Phillips, G.W. & Adcock, E.P. (1996). *Practice Applications of Hierarchical Linear Models to District Evaluations*. Paper presented at American Educational Research Assiciation, 1996 Annual Meeting, NY.

Raudenbush, S.W., & Willms, J.D. (1995). The estimation of School Effects, *Journal of Educational and Behavioral Statistics*, Winter, 20, 4, pp. 307-335.